REES: A Registry of Efficacy and Effectiveness Studies in Education

Dustin Anderson[1], Jessaca Spybrook[1], and Rebecca Maynard[2]

Calls for improving transparency across the social science are increasing. One strategy that is gaining momentum in the quest to increase transparency is the practice of preregistering a study and analysis plan prior to conducting a study. In this article, we examine the potential of preregistration as a strategy for increasing transparency in education studies. We review existing registries within the social sciences and provide a rationale for why we need a registry specifically geared towards education studies. Finally, we introduce the Registry of Efficacy and Effectiveness Studies (REES), developed by the Society for Research on Educational Effectiveness (SREE). The goal of REES is to increase transparency for studies seeking to draw causal conclusions within the education research and evaluation community.

Keywords: decision making; educational policy; evaluation; planning; policy analysis

Openness and transparency have long been recognized as vital for science (e.g., Nosek, Spies, & Motyl, 2012). However, calls to reexamine foundational practices around these norms are becoming prevalent in social science research. The calls for increased transparency stem from concerns that the growing knowledge base may be skewed, incomplete, or untrustworthy (e.g., Kepes, Bennett, & McDaniel, 2014). For example, in a recent systematic review of studies on questionable research practices, 91% of studies found severe evidence of such practices (Banks, Rogelberg, Woznyj, Landis, & Rupp, 2016). Many are concerned with the legitimacy of the knowledge base as a result of some of these questionable research practices (Gehlbach & Robinson, 2018; Ioannidis, 2005; John, Loewenstein, & Prelec, 2012).

In this article, we briefly review practices that may bias the knowledge base that have been documented across the social sciences. Then we examine calls for the adoption and use of publicly available registries and preanalysis plans as a strategy to broaden the scope of accessible research and help mitigate the influences of these practices. Next, we review major registries in the social sciences and provide an argument for why we need an independent registry within education. Then we introduce the Registry of Efficacy and Effectiveness Studies (REES) (https://www.sreereg.org), developed by the Society for Research on Educational Effectiveness (SREE), to increase transparency for studies seeking to draw causal conclusions within the education research community.[1] We conclude by exploring the role of journals and funders in incentivizing researchers to preregister impact studies in REES. We hope this article highlights the importance of increasing transparency in the education research community and how the implementation of REES has the potential to help achieve this goal.

Reporting Bias

The Cochrane Handbook defines reporting bias as the systematic error associated with reported and unreported findings (Higgins, Altman, & Sterne, 2011). Reporting bias may result from choices researchers make (a) while conducting a study or (b) during the dissemination stage. We briefly review practices in each phase that may contribute to reporting bias and the factors that may inflate these practices. Note that we focus strictly on studies seeking causal conclusions or studies testing the impact of an intervention.

Researchers are faced with many choices while conducting a study that may lead to the manipulation of effect sizes (Miguel et al., 2014). These choices are particularly apparent during the design phase and the data analysis phase. In the absence of prespecified design plans or a priori decision rules, there are

[1]Western Michigan University, Kalamazoo, MI
[2]University of Pennsylvania, Philadelphia, PA

Educational Researcher, Vol. 48 No. 1, pp. 45–50
DOI: 10.3102/0013189X18810513

opportunities for researchers to engage in practices that are problematic. For example, researchers may need to decide what groups to compare if there are more than two conditions, what observations to exclude, what outcome measures to analyze, and so forth. If they make these decisions on the basis of obtaining statistically significant findings, often known as p-hacking, they may produce results that appear to be more favorable than they actually are. P-hacking has been attributed to distortions in the distribution of *p*-values in the published literature (Simonsohn, Nelson, & Simmons, 2014). A second and more extreme practice that can lead to bias in the knowledge base is the fabrication of data, which involves intentionally creating new, false data or modifying existing data to produce statistically significant results (National Academy of Sciences, 1992). Instances of this are rare but do exist in the literature base (Fanelli, 2009).

Practices that occur while reporting findings from a study, or during the dissemination process, can also contribute to bias in the knowledge base. Selective outcome reporting, HARKing, and the file drawer problem are three well-known reporting practices that lead to publication bias. Selective outcome reporting occurs when only a select number of outcomes measured in a study are reported as opposed to reporting on the full set of outcomes that were measured (Norris et al., 2013). Typically, the outcomes that are reported are those that are statistically significant. HARKing (Kerr, 1998) refers to an instance when a researcher presents a post hoc hypothesis as an a priori hypothesis or, similarly, presents exploratory results as if they are confirmatory results. Often these post hoc hypotheses are those that resulted in statistically significant findings. The file drawer problem refers to the lack of null findings being reported and/or published and can lead to an increase of positive results in published literature (Franco, Malhotra, & Simonovits, 2014).

The dissemination process may also be influenced by external structures/pressures that contribute to a biased knowledge base. Studies have revealed evidence of journals' propensity to favor studies that show statistically significant positive effects for publication (Franco et al., 2014). This trend has been particularly evident in top-tier journals, as they are in a competitive market and seek to publish studies that would be highly cited (Gerber, Malhotra, Dowling, & Doherty, 2010). Concurrent with journals favoring significant effects, there are often external incentives and pressure on researchers to publish (Leis-Newman, 2011; Pigott, Valentine, Polanin, Williams, & Canada, 2013). Faculty positions specifically incentivize publication through employment actions such as tenure and promotion, which also carry financial benefits (Brodeur, Lé, Sangier, & Zylberg, 2013; Gerber et al., 2010). Further, in a world where the employment of many researchers rests on the acquisition of external grant funding, statistically significant results may also affect further funding opportunities. That is, publishing positive, statistically significant findings may positively influence one's career advancement or help increase the chances of funding for a follow-up study or a different study. These external structures and systems might attract researchers to use methods that increase the probability of publication (John, Loewenstein, & Prelec, 2012; Nosek et al., 2012).

Registries and Preanalysis Plans

In an effort to minimize practices that may contribute to reporting bias and increase transparency, we have started to see an increase in attention to publicly available registries of studies (Miguel et al., 2014). A registry is a public database where researchers register their studies before the study begins, during the study, or upon completion (Banks & McDaniel, 2011). If registries are searchable by intervention and study characteristics, such registration of studies can help mitigate the effects of publication bias within the knowledge base by increasing researcher access to information on all studies, including those with findings that are not statistically significant, which may not be published and otherwise be difficult to find (Casey, Glennerster, & Miguel, 2012; Ioannidis, Munafò, Fusar-Poli, Nosek, & David, 2014).

A basic registry may include information such as name of principal investigator, funder, dates of study, and so forth. A more in-depth prospective registration might also include a preanalysis plan (PAP), or a prestudy plan that explicates details of the analysis protocol including but not limited to planned primary and secondary outcome variables, outcome measures, covariates, and/or plans to handle missing data or multiple comparisons in the same analytic domain (Gelman & Loken, 2013; Olken, 2015).

A PAP allows the prespecified plans and any subsequent post hoc exploratory analyses to be distinguishable (Ioannidis et al., 2014). In addition, it minimizes the researcher's flexibility around analysis of confirmatory research questions, which reduces the likelihood of p-hacking or fabrication (Brodeur et al., 2013). In essence, specification of a PAP increases the confidence in the findings (Miguel et al., 2014; Olken, 2015).

Current Registries in the Social Sciences

The calls for transparency have led to the launch of various registries across the social sciences. Currently, there are four primary registries in the social sciences including (a) the American Economic Association's registry of Randomized Control Trials known as the AEA RCT Registry, (b) the International Initiative for Impact Evaluation's Registry for International Development Impact Evaluations (RIDIE), (c) the Evidence in Governance and Politics (EGAP) registry, and (d) the Open Science Framework (OSF) Registry launched by the Center for Open Science (see Table 1). Although each of these four registries seeks to promote information sharing and increase transparency and accountability, we assert that the education research community will benefit from a stand-alone, independent registry, much like the field of medicine. Our rationale is based on the theory that a registry will be easier to use and more useful to the larger community if (a) it has a relevant, targeted substantive focus, (b) it includes all pertinent designs for the substantive field, and (c) it allows for easy and efficient searching and exporting of relevant studies. As we describe next, while each of these four registries has strengths, none of the four registries in Table 1 meet all three of these criteria for education impact studies.

Table 1
Basic Information on the Four Primary Existing Registries in the Social Sciences

	American Economic Association Registry of Randomized Control Trials (AEA RCT Registry)	Registry for International Development Impact Evaluations (RIDIE)	Evidence in Governance and Politics Registry (EGAP)	Open Science Framework Registry (OSF)[a]
Sponsoring group	American Economic Association	International Initiative for Impact Evaluation (3ie)	Evidence in Governance and Politics	Center for Open Science
Substantive focus	Economics, political science, and other social sciences	International development	Governance and politics	Any topic area
Types of study	Impact studies	Impact studies	All types of studies	All types of studies
Designs[b]	RCT	RCT RDD Matching Dif in dif/FE Natural experiment IV Regression with controls Other	Experiments Field experiments Lab experiments Mixed method Statistics Survey methodology	Experiments Observational study Meta-analysis Other
Website	https://www.socialscienceregistry.org/	http://www.ridie.org/	http://egap.org/content/registration	https://osf.io/registries/#!

[a]There are multiple registry types available within the OSF Registry such as AsPredicted Preregistration, Election Research Preacceptance Competition, and so forth. The information in this table is based on the Prereg Challenge, the OSF Registration type that is the most comprehensive for preregistration of causal impact studies.
[b]Design names are RCT = randomized controlled trial; RDD = regression discontinuity design; IV = instrumental variable; Dif in dif/FE = difference in differences, fixed effects.

We first examine the adequacy of RIDIE and EGAP against these criteria. RIDIE and EGAP are focused on studies related to international development and governance/politics, respectively. Researchers searching for impact studies in the domains of international development and government/politics, respectively, would likely find these registries very useful. However, an education researcher searching for impact studies would not likely search RIDIE or EGAP since education is not the substantive focus of these registries. Hence, it would be unlikely that registering an education impact study with either of these registries would make the study more visible within the education research community. This lack of substantive relevance makes RIDIE and EGAP weak fits for education impact studies.

Next we consider the AEA RCT Registry. One could argue that the AEA RCT Registry has a targeted substantive focus, the social sciences; hence, education falls within that domain. Following that line of reasoning, we consider the second criteria: All relevant design options are available. From Table 1, we can see that the AEA RCT Registry is limited to researchers planning a randomized controlled trial (RCT). In education impact studies, we see RCTs, or studies in which units are randomly assigned to condition; quasi-experimental designs (QEDs), or studies in which the treatment and comparison groups are not formed by random assignment; and single case designs (SCDs), or studies with an experiment where an outcome is measured multiple times across various phases, which are defined by whether or not an intervention is present. In fact, all three of these designs are deemed as eligible designs for assessing the impact of an intervention by the What Works Clearinghouse (https://ies.ed.gov/ncee/wwc/).[2] Limiting the registry to only RCTs, as would be the case in the AEA RCT Registry, would exclude many education impact studies, making the AEA RCT Registry too narrow to meet the needs of the education research community.

Finally, we examine the OSF option in light of the three criteria. The OSF Registry, and specifically the Prereg Challenge we highlight in Table 1, includes a broad substantive focus and a wide variety of design options that may be appealing for some researchers. However, the breadth of the registry made it challenging to search and identify a specific set of studies, or education impact studies using an RCT, QED, or SCD, quickly and efficiently (Criteria 3). For example, we conducted a search within the OSF Registries for education impact studies, narrowing results to those registered with the Prereg Challenge. We tried several search strings including *education AND impact, education AND random**, *elementary AND education AND random**. Each search yielded a large number of studies. However, in many cases the titles did not appear relevant. Further, in order to determine whether or not a study was relevant and to learn more about the study details such as grade level, outcome domain, and so forth, a user must click on the study itself and read through each entry, which can be very time-consuming. In terms of export options after a search is conducted, OSF allows a user to view and print individual entries. However, there is no option to export data from multiple studies into a usable format, which can often help a user quickly summarize the findings from the search. As such, we assert that while the breadth of substantive areas and designs in the OSF result in an extensive database, this database can be challenging to navigate and export when there is interest in one particular type of study—in this case, an education impact study.

To illustrate the value of a registry that meets all three criteria, we briefly turn to the field of medicine and clinicaltrials.gov, a

registry that has a clear substantive focus, accommodates relevant designs, and is easily searchable. Clinicaltrials.gov is the central source for researchers conducting clinical trials to preregister their studies. The questions are tailored to clinical trials and use language that is familiar to researchers planning and conducting clinical trials, which makes it easier for researchers to enter their studies. Further, researchers, practitioners, and individuals searching for clinical trials know to go to ClinicalTrials.gov to search for clinical trials and are able to easily search for in-process or completed trials on a given outcome domain (condition or disease) within a relevant sample age group. Just like in medicine, by creating one registry of impact studies for education, we aim to make it easy for those entering studies and those searching for studies. As we discuss in detail in the next section, the targeted substantive focus of REES allows for the use of language that is familiar to education researchers and relevant to the designs of education impact studies in an effort to make the process of entering a study quick and easy. Further, because researchers must describe their studies using a limited set of categorical terms, searching REES is intended to be easy and efficient, and export options include individual registry entries or Excel-based spreadsheets with data from multiple studies.

The REES

SREE, with the support of the Institute of Education Sciences (IES) (R305U150001), developed and launched REES, a registry for impact studies in education. The vision for REES is to be a reliable source for identifying all impact studies in education, including planned, in-process, or completed studies. We define impact studies as those seeking to determine the efficacy or effectiveness of an educational intervention or strategy (Institute of Education Sciences and National Science Foundation, 2013). Consistent with the designs deemed eligible by the What Works Clearinghouse (WWC) Standards Handbook Version 4.0 (2018), REES accepts RCTs, QEDs, Regression Discontinuity Designs (RDD), and SCDs. Both RCTs and QEDs are commonly used in impact studies in education and are considered acceptable designs by the WWC, although the two types of studies differ in terms of their highest potential rating under the WWC Group Design Standards with RCTs having the potential to meet standards without reservations and QEDs having the potential to meet standards with reservations. RDDs are also used in impact studies in education and can meet WWC RDD standards with or without reservations. SCDs have a set of pilot standards through the WWC, and an SCD has the potential to either meet the WWC Pilot SCD standards with or without reservations.

In addition to the goal of trying to increase transparency and potentially reduce reporting bias, we are optimistic that the establishment of REES will improve education research, policy, and practice in several other ways. First, completing a REES entry compels researchers to think carefully about all aspects of the study as a PAP is included in registering a study. We believe this has the potential to improve the overall methodological rigor and quality of the study design and analysis. Second, it allows researchers, policymakers, and funders to easily identify studies that are in process or complete, which we anticipate will make it easier to identify gaps in the research and areas to invest resources. Third, we hope that as the REES database grows, it will facilitate more efficient planning and expedite the process of conducting (a) research syntheses, as studies that are not in the published literature will be more easily located; and (b) replication studies, since study details and PAPs are a part of a REES entry. Fourth, and also contingent on growing the REES database, we believe it has the potential to provide a valuable mechanism for assessing the extent and nature of publication bias in education research.

As noted above, REES was designed specifically to accommodate impact studies in education. In an effort to make the registry accessible to education researchers, the language used is similar to that used by key infrastructures in education such as the WWC, IES, and the National Science Foundation (NSF). Recognizing that studies often undergo changes, REES allows users to update study entries and chronicles study changes in a clear and nonjudgmental manner. Changes are time-stamped, and researchers are encouraged to include a narrative description of the changes.

REES is an interactive website. It was designed with the goal of enabling researchers to quickly and easily create a registry entry. Any designated study administrator can enter study data into REES and make updates at any time in the future. A designated collaborator can view the entry while the study data are entered but cannot make changes to the entry. Registry entries can be started and stopped at any time, and a portable document format (pdf) version of a partially complete or fully complete entry can be saved, downloaded, and printed at any point. The aim is that a study with a detailed proposal, such as an IES-funded Goal 3, efficacy or replication project, should transfer easily into a REES entry. Entries within REES are searchable and can be exported into an Excel file.

A registry entry includes basic study information as well as details related to the design and analysis plan, or the PAP. A complete REES entry includes eight sections and the following information:

- Section 1: General Study Information
    - Study title, principal investigator(s) names and affiliations, registration date, funder(s), award number, institutional review board (IRB) approval date and number, any other registration numbers, study start and end date, intervention start and end date, phase of study, brief abstract, keywords
- Section 2: Description of Study
    - Type of intervention, topic area, number of intervention arms, target school level, target school type, locations of implementation, brief description of intervention condition(s), brief description of comparison condition
- Section 3: Research Questions
    - Description of confirmatory and exploratory research question(s)
- Section 4: Study Design
    - Identification of research design including presence of blocking, unit of assignment, probability of assignment, unit outcome data is measured
- Section 5: Sample Characteristics
    - Number of units in the intervention(s) conditions at each level, number of units in the comparison

condition at each level, sample exclusion criteria at each level, sample inclusion criteria at each level

- Section 6: Outcomes
  - Number of outcome(s) for each confirmatory question, for each outcome—domain, name of outcomes measure, scale associated with outcome measure, whether same outcomes are being collected in both groups
- Section 7: Analysis Plan
  - Description of baseline measures, identification of covariates to be included in the model at each level, description of analytic model, plan for missing data
- Section 8: Additional Materials
  - Links to study data, links to reports or study websites, links to publications, or upload of relevant files such as study proposals, findings, data, measures, and so forth

A unique feature of REES is the manner in which information is captured. To the extent possible, the information is collected through questions with discrete response categories. This serves three purposes. First, it promotes consistency in language across REES entries. For example, in a narrative description of the design, a user might talk about how clusters are randomly assigned to conditions and call the design a cluster randomized trial. A different user may call the same design a group randomized trial or a field trial. In REES, once a user selects RCT as the design option, she or he answers a series of questions to identify the specific type of RCT, resulting in the same design names being used across all entries. Second, it allows users to more easily search the database by study characteristics such as design, topic, or grade level since responses are recorded and stored using the same categories for all studies. Third, it ensures a minimum depth of information for each study. For example, in a narrative form, the information related to the primary outcomes for a study could vary greatly across studies. In REES, a standard set of information, such as the number of outcome measures per research question and the outcome domain for each research question, is elicited using discrete responses for all studies. Additional information can be included in Section 8 of the registry, but completing Sections 1 through 7 should provide a similar level of detail for all studies.

Engaging the Research Community

REES was launched in October 2018. Engagement by the research community is critical in the success of REES. This includes funders, journal editors, and researchers. Funders of education research play a key role in the likelihood researchers will preregister impact studies in education in REES. The role of funders could vary from encouraging researchers to preregister studies in REES to mandating preregistration. As an example, EGAP's Metaketa initiative mandates preregistration of analysis details prior to collecting outcome data in order to obtain funding (Dunning, 2016). In education, the Field Year 2019 IES Request for Applications (RFA) (https://ies.ed.gov/funding/ncer_progs.asp) states that Data Management Plans for Goal 3 studies should include "Plan for pre-registering the study in an education repository (e.g. see the SREE Registry of Efficacy and Effectiveness Studies)" (p. 79). This represents the first year that plans for preregistration were included as part of the RFA. Journal editors also play a key role in how they support preregistration. At the extreme end, journals may require preregistration for publication. For example, in medicine, the International Committee of Medical Journal Editors released a statement that studies must be preregistered in order to be considered for publication (De Angelis et al., 2005). For journals not yet ready to require preregistration, badges of transparency for reported results that were pre-registered may be added to publications. We have seen examples of this in some areas of the social sciences including *Psychological Science*, which gives authors an opportunity to earn transparency badges if they meet established criteria (Kidwell et al., 2016). It is important to note new funding or publication policies around preregistration would require structural changes as each would need qualified reviewers in charge of verifying such preregistration. Lastly, support for preregistration from researchers planning impact studies is critical. This support may come in the form of not only preregistering their own studies but also encouraging others to preregister relevant studies and embracing the push towards more transparency to improve education research. Like other areas of the social sciences, we are at a critical time, and we believe that active engagement and participation in REES across the education research community has the potential to improve the rigor and credibility of education research in the future.

NOTES

Dustin Anderson and Jessaca Spybrook are joint first authors.

[1]The authors of this article led the design and development of REES. REES was supported by a grant from the Institute of Education Sciences (R305U150001).

[2]Note that single-case design standards and procedures are still in the pilot phase.

REFERENCES

Banks, G. C., & McDaniel, M. A. (2011). The kryptonite of evidence-based I–O psychology. *Industrial and Organizational Psychology*, *4*(1), 40–44. https://doi.org/10.1111/j.1754-9434.2010.01292.x

Banks, G. C., Rogelberg, S. G., Woznyj, H. M., Landis, R. S., & Rupp, D. E. (2016, September). Editorial: Evidence on questionable research practices: The good, the bad, and the ugly. *Journal of Business and Psychology*, 31(3), 323–338.

Brodeur, A., Lé, M., Sangier, M., & Zylberg, Y. (2013). *Star wars: The empirics strike back* (IZA Discussion Paper No. 7268). Bonn, Germany: Institute for the Study of Labor (IZA).

Casey, K., Glennerster, R., & Miguel, E. (2012). Reshaping institutions: Evidence on aid impacts using a pre-analysis plan. *Quarterly Journal of Economics*, *127*(4), 1755–1812.

De Angelis, C. D., Drazen, J. M., Frizelle, F. A., Haug, C., Hoey, J., Horton, R., . . . International Committee of Medical Journal Editors. (2005). Is this clinical trial fully registered?—A statement from the International Committee of Medical Journal Editors. *New England Journal of Medicine*, *352*(23), 2436–2438. https://doi.org/10.1056/NEJMe058127

Dunning, T. (2016). Transparency, replication, and cumulative learning: What experiments alone cannot achieve. *Annual Review of Political Science*, *19*(1), S1–S23. Retrieved from https://doi.org/10.1146/annurev-polisci-072516-014127

Fanelli, D. (2009, May 29). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS one*, *4*(5), e5738. https://doi.org/10.1371/journal.pone.0005738

Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, *345*(6203), 1502–1505.

Gehlbach, H., & Robinson, C. D. (2018). Mitigating illusory results through preregistration in education. *Journal of Research on Educational Effectiveness*, *11*(2), 296–315.

Gelman, A., & Loken, E. (2013). *The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time*. Retrieved from http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf

Gerber, A., Malhotra, N., Dowling, C. M., & Doherty, D. (2010). Publication bias in two political behavior literatures. *American Politics Research*, *38*(4), 591–613. doi:http://dx.doi.org/10.1561/100.00008024

Higgins, J. P., Altman, D. G., & Sterne, J. A. (2011). Chapter 8: Assessing risk of bias in included studies. In J. P. Higgins & S. Green (Eds.), *Cochrane handbook for systematic reviews of interventions* (5.1.0 ed.). The Cochrane Collaboration. Retrieved from www.handbook.cochrane.org

Institute of Education Sciences and National Science Foundation. (2013). Common guidelines for education research and development. Retrieved from http://ies.ed.gov/pdf/CommonGuidelines.pdf

Ioannidis, J. P. (2005, August). Why most published research findings are false. *PLoS Med*, *2*(8), 0696–0701.

Ioannidis, J. P., Munafò, M. R., Fusar-Poli, P., Nosek, B. A., & David, S. P. (2014). Publication and other reporting biases in cognitive sciences: detection, prevalence and prevention. *Trends in Cognitive Science*, *18*(5), 235–241. Retrieved from http://doi.org/10.1016/j.tics.2014.02.010

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23*(5), 524–532.

Kepes, S., Bennett, A. A., & McDaniel, M. A. (2014). Evidence-based management and the trustworthiness of our cumulative scientific knowledge: Implications for teaching, research, and practice. *Academy of Management Learning & Education*, *13*, 446–466. Retrieved from http://dx.doi.org/10.5465/amle.2013.0193

Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, *2*, 196–217.

Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L.-S., . . . Nosek, B. A. (2016). Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency. *PLOS Biology*, *14*(5), e1002456. https://doi.org/10.1371/journal.pbio.1002456

Leis-Newman, E. (2011). Securing tenure: On the tenure track? Here are four keys to making sure you get the ultimate prize. *Monitor on Psychology*, *42*(5), 7.

Miguel, E., Camerer, C., Casey, K., Cohen, J., Esterling, K. M., Gerber, A., . . . Van der Lann, M. (2014). Promoting transparency in social science research. *Science*, *343*(30), 30–31.

National Academy of Sciences. (1992). *Responsible science: Vol. I. Ensuring the integrity of the research process*. Washington DC: National Academy Press. Retrieved from https://www.nap.edu/catalog/1864/responsible-science-volume-i-ensuring-the-integrity-of-the-research

Norris, S. L., Moher, D., Reeves, B. C., Shea, B., Loke, Y., Garner, S., . . . Wells, G. (2013, March). Issues relating to selective reporting when including non-randomized studies in systematic reviews on the effects of healthcare interventions. *Research Synthesis Methods*, *4*(1), 36–47. doi:10.1002/jrsm.1062

Nosek, B. A., Spies, J. R., & Motyl, M. (2012, November 7). Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science*, *7*(6), 615–631. doi:10.1177/1745691612459058

Olken, B. A. (2015). Promises and perils of pre-analysis plans. *Journal of Economic Perspectives*, *29*, 61–80.

Pigott, T. D., Valentine, J. C., Polanin, J. R., Williams, R. T., & Canada, D. D. (2013). Outcome-reporting bias in education research. *Educational Researcher*, *42*(8), 424–432.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: A key to the file drawer. *Journal of Experimental Psychology: General*, *143*, 534–547. doi:10.1037/a0033242

What Works Clearinghouse (2018). *Standards handbook* (Version 4.0). Retrieved from https://ies.ed.gov/ncee/wwc/Handbooks

AUTHORS

DUSTIN ANDERSON is an assistant director of research for the High Impact Leadership Project at Western Michigan University, 1903 W. Michigan Ave, Kalamazoo, MI 49008-5283; *dustin.anderson@wmich.edu*. His research focuses on strategies for increasing transparency across education research and increasing leadership capacity within schools that support school renewal initiatives.

JESSACA SPYBROOK is a professor of evaluation, measurement, and research at Western Michigan University, 1903 W. Michigan Ave, Kalamazoo, MI 49008-5283; *Jessaca.spybrook@wmich.edu*. Her research focuses on improving the design of large-scale field trials in education.

REBECCA MAYNARD is the university trustee professor of education and social policy at the University of Pennsylvania, 3700 Walnut Street, Philadelphia, PA 19104; *rmaynard@upenn.edu*. Her research focuses on methods for integrating program evaluation and improvement science and on improving the quality and utility of research syntheses.

Manuscript received December 4, 2017
Revisions received April 11, 2018, and
September 14, 2018
Accepted October 8, 2018